A polymer physics view on universal and sequence-specific aspects of chromosome folding

Daniel Jost*
University Grenoble Alpes, CNRS, TIMC-IMAG lab, UMR 5525, Grenoble, F-38706 La Tronche, France

Angelo Rosa†
SISSA (Scuola Internazionale Superiore di Studi Avanzati), Via Bonomea 265, 34136 Trieste, Italy

Cédric Vaillant‡ and Ralf Everaers§
Univ Lyon, ENS de Lyon, Univ Claude Bernard Lyon 1, CNRS, Laboratoire de Physique and Centre Blaise Pascal, F-69342 Lyon, France

(Dated: September 7, 2016)

Recent advances in genome-wide mapping and imaging techniques have strikingly improved the resolution at which nuclear genome folding can be analyzed and revealed numerous conserved features organizing the one dimensional chromatin fiber into tridimensional nuclear domains. Understanding the underlying mechanisms and the link to gene regulation requires a cross-disciplinary approach that combines the new high-resolution techniques with computational modelling of chromatin and chromosomes. The present chapter discusses our current understanding of generic aspects of chromosome behaviour during interphase. In particular, we present explanations from polymer physics for the emergence of the universal "territorial" folding of chromosomes *above* the Mbp scale and the sequence-dependent formation of *topologically associating domains* (TADs) *below* the Mbp scale.

I. INTRODUCTION

Eukaryotic genomes are partitioned into single, independent functional units, the chromosomes. Each chromosome contains a unique, polymer-like filament of double-helical DNA carrying the genetic information. Its total length can be measured in basepairs (bp) or, more commonly for very long chromosomes, thousands (kilo-basepairs, kbp) or millions of basepairs (mega-basepairs, Mbp). With a total of $\approx 7.0 \times 10^9$ bp split into 2×23 chromosomes, *Homo sapiens* is fairly typical for the estimated $(8.7 \pm 1.3) \times 10^6$ [1] species of eukaryotes currently living on our planet. Chromosome numbers range from 2×4 for *Drosophila* to 2×225 in a butterfly species [2] and reach of the order of 1000 in some polyploid ferns [3]. Genome sizes are between $\approx 2.4 \times 10^7$ bp for budding yeast (*Saccharomyces cerevisiae*) and can reach $\approx 10^{11}$ bp in amphibians and flowering plants [4, 5]. On the other hand, chromosome sizes can be either as small as ~ 0.2 Mbp in *S. cerevisiae* [6] or as large as in the case of the Japanese plant *Paris japonica* where one single chromosome amounts to $\sim 4 \times 10^3$ Mbp [7].

For most of the time of the cell cycle, namely during the so-called *interphase* between cell divisions, the chromosomes of eukaryotic cells are confined to a specialised region, the nucleus [8]. Chromosome folding inside the nucleus is highly variable, but not random [9], and increasingly studied with visual and chromosome capture techniques (see, for instance, Ref. [10–13] and Part III of this book). In the present, brief review, we concentrate on features of interphase chromosomes, which can be understood in terms of the same theoretical and/or computational tools [14–17] which have been successfully applied to synthetic polymers and the related fields in soft matter physics [18–20]. In particular, we present explanations from polymer physics for the emergence of (1) the universal "territorial" [21–23] folding of long chromosomes *above* the Mbp scale; (2) the sequence-dependent formation of *topologically associating domains* (TADs) *below* the Mbp scale [24, 25]. The manuscript is organized as follows: in sections II A and II B we summarise relevant experimental observations, while the theoretical and modelling results are discussed in sections III and IV. We conclude in section V and discuss perspectives for future work.

II. EXPERIMENTAL INSIGHT ON NUCLEAR GENOME ORGANIZATION: FROM DNA TO TADS AND CHROMOSOME TERRITORIES

To fix ideas and introduce the subject, consider the example of the human genome. The nuclei of human cells have a linear diameter of ≈ 10 microns (μm) and contain DNA with a total contour length of about 2 meters [8]. Stretched out to their full extension of a few centimeters, the DNA of individual human chromosomes exceeds the diameter of the nucleus by more than three orders of magnitude. The association with histone proteins leads

*Electronic address: daniel.jost@imag.fr
†Electronic address: anrosa@sissa.it
‡Electronic address: cedric.vaillant@ens-lyon.fr
§Electronic address: ralf.everaers@ens-lyon.fr

to compaction [8]. However, with a contour length in the millimeter range, the resulting *chromatin* fibers are still *strongly* folded when confined to the nucleus [26]. Owing to the refinement of experimental techniques, considerable progress has been made in recent years in investigating the nuclear structure and dynamics. In particular, biologists have now access to positions and distances [9], mobilities [27] and contact probabilities [28] for (pairs of) specific genomic loci. Variations between *different* chromosomes, cell types, species in the spatial organisation of the genome and the response to specific environmental stimuli provide important *specific* insights into structural mechanisms of genome activity regulation. In contrast, the generic (*i.e.*, sequence-independent) chromosome behavior emerges by averaging experimental data over different genomic sequences or between different cells.

A. Universal aspects of chromosome folding

During interphase, chromosomes *de*condense and appear to loose their identity. However, confirming [33] pioneering observations by Carl Rabl [34] dating back to the ending of XIX century, FISH-labelling reveals a remarkable universal motif in nuclear organisation: chromosomes occupy distinct *territories* and do not mix [9, 35]. Disregarding sequence and considering chromosome folding as a function of genomic distance reveals additional interesting features (Fig. 1). In panel a we show a compilation of experimental data for the mean-square spatial distance, $\langle R^2(N_{bp})\rangle$ as a function of the number of base pairs (bp), N_{bp}, separating two sites along a chromosome. Panel b contains *sequence averaged* contact probabilities, $p_c(N_{bp})$. All biological specificity is suppressed: in fact, data for yeast and human chromosomes nicely superpose on corresponding length scales (around 0.1 Mbp). Moreover, and as often in polymer physics, the data exhibit power law behavior with $\langle R^2(N_{bp})\rangle \propto N_{bp}^{2\nu}$ and $p_c(N_{bp}) \propto N_{bp}^{-\gamma}$ characteristic of fractal structures [16]. But unlike textbook [14–16, 36] polymers in concentrated solutions where $\nu = 1/2$ and $\gamma = 3/2$, the large scale behavior of long chromosomes seems to be characterized by "non-canonical" exponents $\nu \approx 1/3$ [37] and $\gamma \approx 1$ [28] (Fig. 1). In section III, we will rationalize these observations in terms of a coherent theoretical framework based on polymer physics.

B. Sequence-specific aspects of chromosome folding

In recent years, genome-wide chromosome conformation capture techniques (HiC [28]) have revealed reproducible, sequence and epigenetic state dependent features in the nuclear organization of chromatin, at the sub-chromosomal ($\lesssim 1$ Mbp) scale (Fig. 2a). They are interpreted as a 3D segmentation into *topologically associating domains* (TADs) characterized by an enrichment of

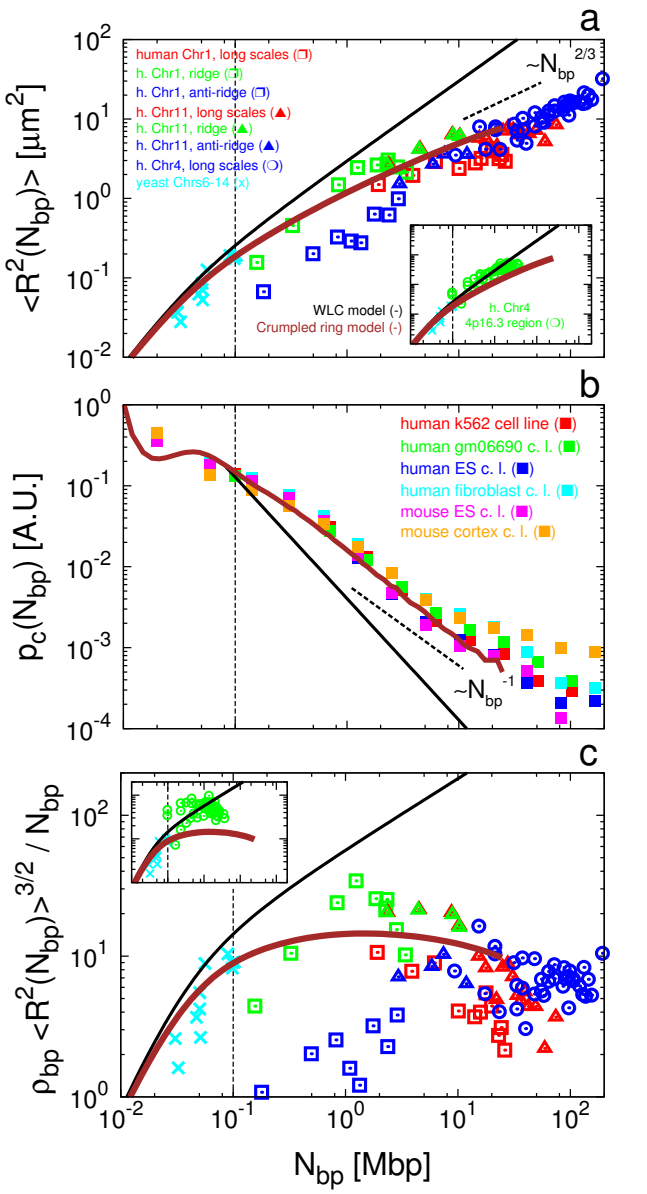

FIG. 1: Experimental behavior of interphase chromosomes (symbols) compared to generic polymer models (solid lines). (a) Mean-square internal distances, $\langle R^2(N_{bp})\rangle$, between genomic sites separated by N_{bp} Mbp along the chromatin fiber: experimental results for interphase yeast [29] and human chromosomes [30, 31] obtained by FISH (symbols), compared to the worm-like chain (WLC) model (black line) and the crumpled ring model [32]. The inset reproduces FISH data from the "equilibrated" 4p16.3 terminal region on human chromosome 4 [22]. The crumpled ring model deviates only from data for the anti-ridge region on human chromosome 1. (b) Average contact probabilities between genomic sites: experimental results for human and mouse chromosomes in different cell lines measured by HiC [24, 28] (symbols), and corresponding predictions for the WLC model and the crumpled ring model (solid lines). (c) Overlap parameters corresponding to the data shown in (a). Taken together, these data are consistent with expected deviations from the ideal WLC behavior (black lines) occurring in the "bulk" of eukaryotic chromosomes when $\frac{\rho_{bp}}{N_e}\langle R^2(N_e)\rangle^{3/2} \equiv 20$, with $N_e \approx 10^5$bp (vertical dashed lines).

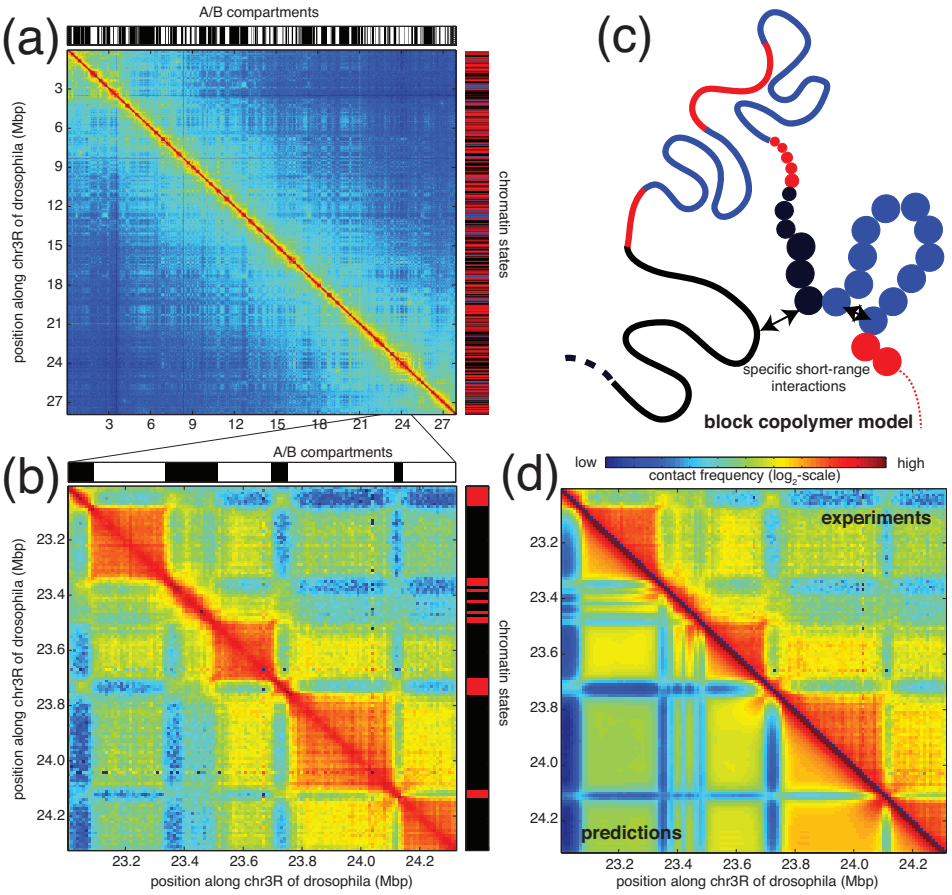

FIG. 2: (a) Contact frequency map across chromosome 3R of *Drosophila melanogaster* based on HiC data [25] at 30 kbp bin size. The checker-board-like pattern of long-range contacts allows the definition of two compartments A and B (black and white segmentation at the top). The local chromatin states [38] are given at the right-hand side (red for active, blue for facultative heterochromatin, green for constitutive heterochromatin and black for null heterochromatin). (b) Zoom on the 23-24.3 Mbp region at 10 kbp bin size showing the local segmentation into TADs. (c) Block copolymer model of chromatin: each monomer is characterized by a chromatin state. Short-range specific interactions are considered between monomers of the same state. (d) Illustration of the predictive outcome of the copolymer model for the region displayed in (b).

intra-domain contact frequencies and a partial 3D insulation between adjacent domains (Fig. 2b) [39, 40]. Their sizes vary from few kbp up to Mbp. TADs are observed in many species ranging from yeast to human [41], and have been shown to be conserved during cellular differentiation [42] and even between close species [40, 43]. At larger scales, HiC maps of higher eukaryotes display a characteristic cell-type specific checker-board-like pattern where TADs engage in long-range interactions (Fig. 2a) [28]. Statistical analysis of the local enrichment (or depletion) of contacts compared to the average behavior has demonstrated the presence of two main compartments (often named A and B) that partition the genome at a higher scale: contacts between genomic regions or TADs belonging to the same compartment are more frequent than between regions of different ones [28, 44]. In general, the A compartment is mainly composed of active – euchromatic – regions while B is more repressed and heterochromatic. These compartments may eventually be subdivided into sub-compartments, characterizing sub-states of the chromatin [25, 43]. This highlights the strong correlation between the global 3D chromatin organization and the local activities or states of the chromatin [25, 38, 45, 46].

A key question concerns the mechanisms behind the formation of TADs and compartments. Again, polymer physics may be a powerful tool to build minimal models for investigating the validity of proposed processes. In section IV, we will discuss the role of specific interactions in *hetero*polymer models, which can selectively stabilize some of the transiently appearing branched loop structures from the generic *homo*polymer models to be discussed in section III.

III. UNIVERSAL ASPECTS OF CHROMOSOME FOLDING: POLYMER THEORY

We begin our analysis with the large-scale, generic features of chromosome folding summarised in section II A.

This choice is not obvious. Physical modelling proceeds from small to large scales and one might be tempted to dismiss the generic features as a "vague echo" of biologically relevant structures, which are defined through contacts between specific genomic sites and which are maintained by a complex, evolved molecular machinery. The fractal nature of the chromosome conformations would then be a mere curiosity. Instead, we adopt (and explain) the opposite point of view that sequence averaging reveals the *generic*, polymer-like structure and dynamics of interphase chromosomes. We show that the available experimental evidence for their behavior can be *quantitatively* predicted by maximizing the entropy of a chromatin fiber model under the constraint that chromosomes are free of knots and not entangled with each other (brown solid lines in Fig. 1). As a consequence, and as largely emphasized at the end of section II A, the emerging picture of the folding of interphase chromosomes departs from the "traditional" one for linear chains in equilibrated solutions or melts [14, 36]. In particular, we believe that the proper modelling of topological constraints and the largely knot-free microscopic topological state of interphase chromosomes prior to replication represent an essential features of models for sequence specific aspects of chromosome folding.

A. Chromatin fiber entanglement

Given the controversial fiber structure [47] and the complexity of chromatin on the molecular scale, it is far from obvious, that polymer physics has relevant qualitative or even quantitative insights to offer. A polymer model characterizes chains by their contour length, L, and their Kuhn length, l_K, as a measure of the chain stiffness. For contour lengths $L \ll l_K$ thermal fluctuation have little effect and the chains are effectively rigid with mean square end-to-end distances $\langle R^2(L)\rangle = L^2$ and $\nu = 1$. For $L \gg l_K$, equilibrated linear chains exhibit random coil statistics with $\langle R^2(L)\rangle = l_K L$ and $\nu = 1/2$. In this regime, the contact probability, p_c, between two segments scales like $p_c(L) \sim (L/l_K)^{-\gamma}$ with $\gamma = 3\nu = 3/2$. The crossover for $L \approx l_K$ can be conveniently described by the worm-like chain (WLC) model [48, 49], excluded volume interactions being screened in concentrated solutions [14]. For 30nm chromatin fibers, $L = 0.01 N_{bp}$ nm and $l_K \approx 300$ nm [29]. For the locally much less compact 10nm fibers, a simple estimate [50] assuming uncorrelated orientations of subsequent nucleosomes yields $L = 0.125 N_{bp}$ nm and $l_K \approx 25$ nm. Interestingly, the two fiber models predict with $\langle R^2(N_{bp})\rangle \approx 3 N_{bp}$ nm^2 identical mean-square internal distances in the random walk regime for genomic distances larger than ≈ 30 kbp, suggesting that this estimate should be relatively robust and even apply to fibers, whose local structure alternates between dense 30nm and open 10nm conformations [51]. As shown in Figs. 1a and 1b, the predictions of the WLC model (black lines) are in reasonable agreement with the sequence-, cell-type and species-averaged experimental data for genomic distances up to ≈ 100 kbp and, in exceptional cases such as equilibrated telomeric regions [22], even on the Mbp scale.

Similarly to macroscopic strings tied into knots, diffusing polymer chains can slide past each other, but their backbones cannot cross. The resulting topological constraints [52, 53] start to affect polymers beyond the so called entanglement length, N_e [14]. According to the packing argument for loosely entangled chains [54–56], N_e can be determined from the condition that the so-called overlap parameter, $\Omega(N_{bp}) \equiv \frac{\rho_{bp}}{N_{bp}} \langle R^2(N_{bp})\rangle^{3/2}$, reaches a characteristic threshold, $\Omega \equiv 20$ [32, 54–56]. For typical nuclear densities of $\rho_{bp} \approx 0.011$ bp/nm^3 both fiber models suggests an entanglement length for genomic DNA of the order of [22]

$$N_e = 1.2 \times 10^5 \text{bp}. \qquad (1)$$

Note, however, that this crucial length scale is strongly density dependent [19, 56]. The relevance of topological constraints for the structure of chromosomes can be verified directly from the experimental data. In panel 1c we have plotted dimensionless packing ratios inferred from FISH data. Comparison with panels 1a and b shows that deviations from the worm-like chain behavior set in on length scales, where the overlap parameter approaches the entanglement threshold of 20. Qualitatively, the constant overlap parameter on large scales is compatible with a $\nu \approx 1/3$ regime, where the chain extension is controlled by the entanglement threshold. The corresponding time scale for the onset of entanglement effects, $\tau_e \approx 32$s, can be estimated [22] by reinterpreting the anomalous diffusion of a fluorescently-labeled site [57] in terms of the characteristic slowing down of the polymer motion on the entanglement scale.

B. Chromosome conformations as crumpled, randomly branched ring polymers in solution

For *linear* chains, topological constraints are transient. Typically, they dominate the viscoelastic behavior of long-chain melts or solutions [14, 58], but do not affect the *equilibrium* statistics as the systems remain ergodic. However, this may not be taken for granted in the case of chromosomes. With entanglement times, τ_e, in the range of minutes and an effective size of $Z = N_{bp}/N_e = 1000$ entanglements, equilibration of the microscopic *topological* state via reptation [59] is expected [22, 60] to require centuries as $\tau_{max} = Z^3 \tau_e$. As a consequence, the topological state of interphase chromosomes prior to replication is not random but *identical* to the topological state during the preceding metaphase step of the cell cycle. In particular, there are no topological *links* different chromosomes.

Grosberg *et al.* [21] were the first to argue along these lines, that chromosomes should be in an essentially unknotted state to perform their function. In particular,

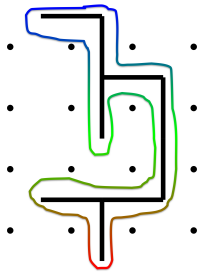

FIG. 3: Illustration of a randomly branched ("lattice tree"-like [32]) ring conformation (rainbow-colored line) with topological constraints ideally represented as an array of fixed obstacles (black dots, see Refs. [59, 65]). In reality, topological constraints are not permanent as they are constituted by surrounding rings which are all subjected to the same stochastic Brownian motion.

they suggested that due to topological constraint chromosomes should fold and interpenetrate differently from polymers in equilibrated melts or semi-dilute solutions. To describe such conformations, they drew an analogy to crumpled globules resulting from the rapid collapse of an isolated polymer chain, which initially preserve the (nearly) unknotted topological state of the good solvent conformation [61]. Recently, this view received strong support from the interpretation of their HiC data by Lieberman-Aiden *et al.* [28], even though the analogy, when taken too literally, does not seem to lead to well-defined structures [62].

As an alternative, two of us (RE, AR) considered [22] the opposite process of *de*condensing initially unknotted and spatially separated (and hence topologically unlinked) metaphase chromosomes in solutions with concentrations corresponding to interphase nuclei. Using a carefully mapped, parameter-free model of chromatin fibers, we were able to reproduce the experimental data [22, 63]. In particular, we were able to show that the bulk of our linear model chromosomes exhibited the same behavior as corresponding *equilibrated*, semi-dilute solutions of *un*entangled *ring* polymers, which show the same "territorial" behavior as interphase nuclei [64].

Understanding this behavior has been a long-standing problem in polymer physics [18, 32, 61, 64, 66–75]. Khokhlov and Nechaev [66] and Rubinstein [69] were the first to argue that such rings should adopt randomly branched, doubled folded conformations, which reduce the threadable surface they present to each other. Fig. 3 illustrates the notion of topologically constrained, randomly branched ring conformations. In a recent study [32], two of us (RE, AR) have validated this idea by developing it into a quantitative multi-scale approach, where a computationally efficient Monte Carlo method is used to generate branched polymer conformations [76], which are subsequently "fine-grained" to corresponding off-lattice conformations of *non-concatenated and unknotted* rings for the fiber model. While the generated

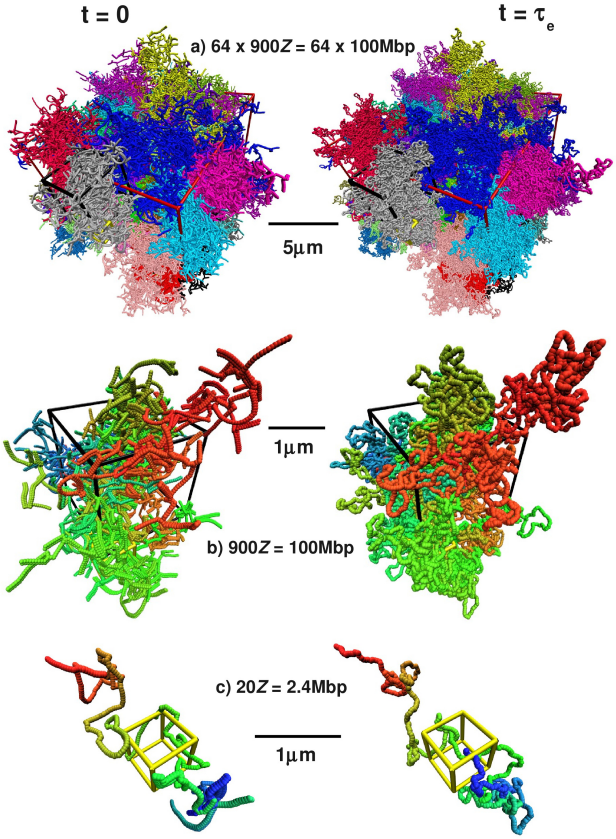

FIG. 4: (a) Model conformations of 64 *interacting* ring polymers, described by the lattice tree model with excluded volume interactions [32]. The contour length of each ring is $N_r = 108$ Mbp or $Z_r \equiv N_r/N_e = 900$, corresponding to the typical size of a human (mammalian) chromosome. (b) Single ring conformation. (c) Ring portion from the single ring conformation in (b), corresponding to $Z = 40$. Boxes indicate the volume, $\left(\frac{\rho_{bp}}{N_r}\right)^{-1}$, available to corresponding configurations at the nominal chromatin density, $\rho_{bp} = 0.011\text{bp/nm}^3$ [22].

conformations are in excellent agreement with the results of brute-force equilibration for $Z \sim 100$, the multi-scale approach provides access to much larger system sizes.

As in our original study [22], all results can be quantitatively mapped to experimental data for chromatin. With $M = 64$ rings of length $Z_r = 900$ our largest systems are comparable in size to the nucleus of a human cell (Fig. 4). Fig. 4a illustrates the characteristic segregation of ring polymers and qualitatively reproduces [22, 64] chromosome territories [9]. Remarkably, Fig. 1 demonstrates (brown lines) that our parameter-free model quantitatively reproduces the available FISH [29, 30] and conformation capture data [24, 28]. Similarly, the reported aspect ratios of chromosome territories of 4.5 : 2.9 : 1.0 [77] closely agree with asymptotic values of 4.9 : 1.9 : 1.0 from the (interacting) lattice tree model. The effective exponents $\nu = 0.32 \pm 0.01$ and $\gamma = 1.11 \pm 0.01$ we

observed in this regime agree with the reported behavior of interphase chromosomes.

Two length scales emerge. First, branching sets in on the entanglement scale of $\sim$ 100 kbp. Interestingly, this is in excellent agreement with the average size of chromatin loops regulated during meiosis [78], and with typical genomic distances between enhancers and corresponding transcription start sites, as reported in a recent study employing chromosome conformation capture techniques [79]. Second, the structures become locally compact (Fig. 1a) on the scale of $\sim$ 1 Mbp or $Z \sim 10$ entanglements, i.e. on the TAD scale. Figure 4b and c illustrate that the territorial segregation also persists inside chromosomes down to this scale, but not below. As anticipated by Grosberg *et al.* [21, 61], our results for ring polymers [32] are relevant to linear chromosomes due to a separation of time scales [22] illustrated in Fig. 5: the times scale for the relaxation of the microscopic topological state of human chromosomes (of the order of centuries, estimated as $\tau_e Z_{chr}^3$ [22, 59, 60] as a function of the *total* chromosome size $Z_{chr} = \mathcal{O}(10^3)$ or $N_{chr} = 10^8$bp) far exceeds the time required for the *structural* relaxation of a topologically constrained chromosome on local scales $Z \leq Z_{chr}$ (estimated as $\tau_e Z^{5/2}$ in the ideal lattice tree regime [65] and estimated from relaxation times in Molecular Dynamics simulations of our fiber model). In particular, we find that during the typical length of a cell cycle of $\approx$ 24hours [8] the local equilibration of chromosomes structure should also proceed up to the $\approx$Mbp scale.

To summarise, the structures emerging from topological constraints in non-concatenated ring melts share many generic features of interphase chromosomes. The chains may be said to be *crumpled* [21, 61], to exhibit a form of random looping [31, 80], and to segregate in subcompartments [37]. There are similarities to the crumpled or fractal globule model of chromosomes [21, 28, 32], but also important differences in that the absence of surface tension in the many-chain system leads to strongly interpenetrating, aspherical territories. As these phenomena spontaneously *emerge* in suitable polymer models [22], the approach can *explain* rather than *describe* generic features of interphase chromosomes, *quantitatively predict* the emerging characteristic length scales, and be *integrated* into more detailed models addressing sequence-specific aspects of chromosome folding.

IV. SEQUENCE SPECIFIC ASPECTS OF CHROMOSOME FOLDING: POLYMER THEORY

Our ability to predict the sequence-averaged structure suggests that we have reached a quantitative, physical understanding of *one* important aspect of chromosome folding. However, the discussion presented in section II B clearly points out that there are many aspects in chromosome biology, which are intimately connected to the DNA sequence. Motivated by the observed correla-

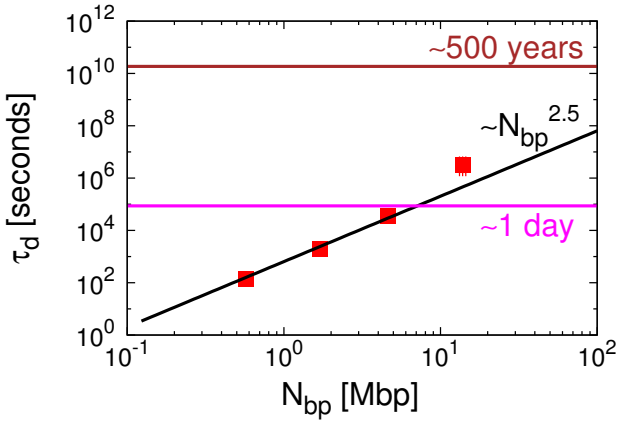

FIG. 5: Time scales separation between equilibration times of ring polymers (red symbols, well described by the $N_{bp}^{5/2}$-power law behavior predicted by Obukhov *et al.* [65]) and the estimated [22] 500-years reptation time of long, linear polymer chains the size the human chromosomes ($N_{bp} \approx$ 100Mbp, brown line). 1 day (magenta line) is the typical time-scale of the cell cycle for most animal cells [8].

tions between the 1D chromatin states and the 3D chromatin organization, heteropolymer models have started to emerge which explicitly consider the coupling between chromatin structure and function [81–94]. These models posit that chromatin folding might likely be driven by direct or effective specific short-range interactions between genomic loci. While the existence of effective interactions in heterogeneous polymers is well established [95], the microscopic foundations of these interactions are still unclear but, in the case of chromatin, may originate from (1) local direct chromatin-chromatin interactions mediated by chromatin-binding proteins with sequence- or epigenetic-specific affinities [96–98] ("block copolymer" model [87] or "binder" models [81, 83]; (2) chromatin fibers with different local packing ratios ("10nm/30nm"-mixed-fibers model [51]) depending for example on epigenetics or gene activity [99]; (3) non-thermal active (ATP-consuming) processes like transcription or chromatin remodelling [100] ("activity-based segregation" model [86]).

In the following, we are going to focus on some recent ideas [87, 101] concerning the connection between polymer physics and the formation of sub-Mbp domains (TADs) inside chromosome territories. Chromatin is modeled as a block copolymer where blocks corresponds to consecutive monomers with an identical chromatin state (Fig. 2c). The dynamics of the chain is then controlled by thermal fluctuations, excluded volume, eventually bending rigidity of the fiber,and attractive short-range interactions between monomers of the same state.

In *Drosophila melanogaster*, two of us (CV, DJ) have extensively studied the behavior of such model at the Mbp scale [87, 101] (Fig. 2d). As explained in sec-

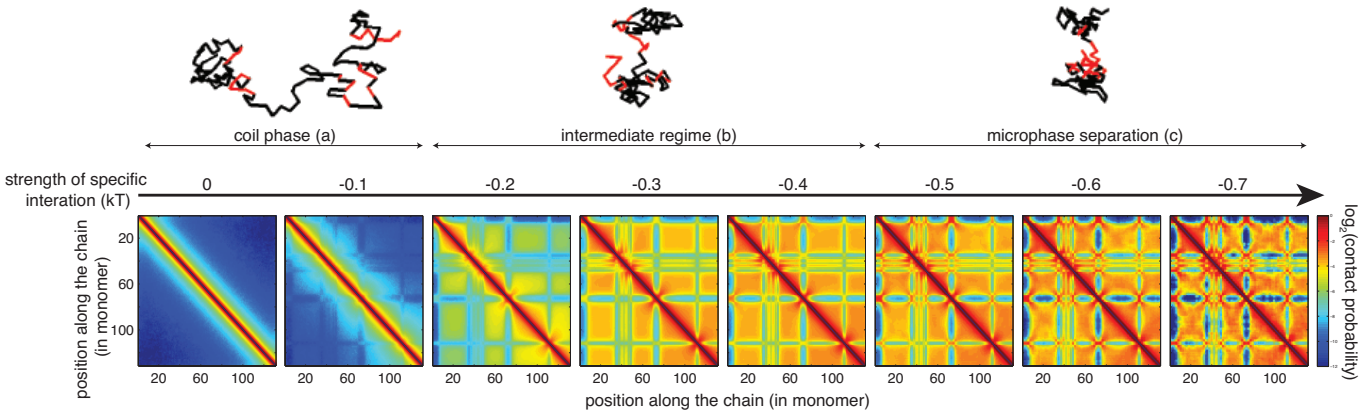

FIG. 6: Contact frequency maps predicted by the copolymer model for the genomic region of *Drosophila* displayed in Fig. 2b, obtained by varying gradually the strength of specific interactions between monomers of the same state. System varies continuously from an unstructured, coil phase (a) to a microphase separation (c), exhibiting an intermediate regime consistent with HiC experiments (Fig. 2d). For each phase, snapshots of typical configurations are shown. Results were obtained using kinetic Monte-Carlo simulations of a model polymer on lattice (see [101]).

tion III, systems can structurally equilibrate at this scale and we neglected topological constraints in the crossover regime to territorial behaviour. Numerical investigations of the block copolymer were performed using either standard Molecular dynamics or kinetic Monte-Carlo simulations or an efficient self-consistent Gaussian approximation [87, 102]. The qualitative behavior of the system is independent of the chosen method. By varying the strength of specific interactions, the systems exhibit a variety of different phases (Fig. 6). For weak interactions, configurations are characteristic of an unstructured, coil phase (Fig. 6a). For strong attractive interactions, a microphase separation is observed and large portions of monomers of the same state occupied separate spatial compartments leading to strong checker-board patterns (Fig. 6c). In the intermediate regime, the systems show a continuous crossover between the coil and the microphase regimes (Fig. 6b). We observe the partial internal collapse of blocks into TAD-like domains, followed by the appearance of weak long-range stochastic interactions between TADs of the same chromatin state. The corresponding 3D compartments may contain several TADs but are transient and only weakly collapsed. As the interactions become more attractive, the blocks experience an internal θ-collapse transition to an equilibrium globule and long-range interactions become more and more important, leading to the formation of long-lived larger 3D compartments. The precise shape of the phase diagram, as well as the behavior of individual blocks, are strongly dependent on the underlying pattern of chromatin states (size of blocks, number of different states, etc.) [87, 101]. For example, larger blocks will start collapsing at weaker interaction strength due to stronger collective effects [101].

Experimental HiC data, with their evidence of the formation of TADs and A/B compartments, are compatible with the intermediate regime (Fig. 2d) where chromatin blocks have partially collapsed into TADs and where blocks of the same state transiently merge together into dynamic 3D compartments resulting in the characteristic weak checker-board pattern of A and B compartments observed in HiC maps. This observation is consistent with FISH microscopy experiments of Polycomb bodies, spatial compartments associated with facultative heterochromatin, showing that such bodies are indeed highly dynamic inside the fly nucleus [103]. In this intermediate regime, prediction of the time-evolution of the contact maps shows that TADs form quickly first, followed by the slow formation of long-range interaction. This is again in agreement with HiC data on synchronized cells along the cell cycle [104]. Another property of systems in this regime is the internal compaction of TADs that increases with the TAD size for a given interaction strength. In *Drosophila*, this simple prediction agrees nicely with the measurements on heterochromatic TADs [101, 105]. Interestingly, for active – euchromatic – domains, the compaction does not depend on the size, again pointing out that active chromatin only weakly interacts with itself. This may reflect a distinct local mode of interaction between chromatin types: active chromatin rather organizes locally via pairwise short-range bridging between discrete specific genomic sites while heterochromatin may interact more continuously via clustering of multiple chromatin loci. This is consistent with more homogeneous internal contact patterns observed for inactive domain and more complex interactome profiles for active domains [106].

V. DISCUSSION AND CONCLUSIONS

In this article, we have summarized the results of our collective efforts to understand chromosome folding in terms of polymer physics. In particular, we have discussed the physical origin of

1. The experimentally observed territorial (section II A) chromosome structure. In our framework, universal aspects of chromatin folding may be understood by the thermal (Brownian) relaxation of topologically-constrained chromatin fibers. Nuclei resemble solutions of densely packed unknotted and unentangled ring polymers which form highly-branched conformations (section III).

2. The formation and structure of so-called *topologically associating domains* (TADs). Here, they arise as the consequence of the self-organization and micro-phase separation of chromatin clusters growing inside a model copolymer with sequence-specific chromatin-chromatin interactions (section IV). The model reproduces with remarkable accuracy the check-board pattern of contact matrices from HiC experiments in *Drosophila melanogaster* (section II B).

While the reported agreement with available experimental data is very encouraging, the two proposed approaches do not pretend to be exhaustive or give a complete explanation to chromosome structure. What is currently missing, which should be also considered as a promising direction for future work?

First, to what concern the *large-scale* ($\gtrsim 1$ Mbp) structure of chromosomes, we should ask if our computational approach is pertinent in the presence of intra- and inter-chromosomal contacts, of confinement by and attachment to the nuclear membrane and matrix [107], or of transcriptional activity. Is it really adequate, to either neglect these features or to view them not as being designed to cause looping, but as stabilizing the large scale conformation of chromatin fibers, which generically adopt fluctuating branched loop structures? The topological constraints lead to the confinement of chromosomes to territories, which are one order of magnitude *smaller* than the nucleus. This key aspect is thus properly represented in bulk studies at the nuclear density. Neglecting confinement is nevertheless an approximation. In nuclei with a few dozen chromosomes, none is very far from the nuclear membrane, even though this finite size effect should be less critical for the chromosome structure on smaller scales. Concerning transcriptional activity, the generic structure and the absence of long-lived entanglements strikes us as a *prerequisite* for the activity (and evolution) of transcription factories [108], rather than a consequence [109] resembling self-organized active structures in the cytoskeleton [110]. We want to stress here, that by all this we do *not* mean to imply that the nuclear architecture of biological organisms can be understood neglecting transcriptional activity, confinement by and attachment to the nuclear membrane and matrix, intra- and inter-chromosomal contacts and, in particular, the evolved *specificity* distinguishing organisms and cell lines. We rather propose to view them as *stabilizing* the large scale conformation of dynamically branched loop structures of chromatin fibers rather than as having evolved *to create* looped equilibrium structures in linear chains in an origami-like [111] fashion. In this respect, we suggest then that some care should be required in addressing the role of specific interactions between different genomic sites or linking chromosomes at designated points to the nuclear membrane and or a nuclear matrix [107].

To conclude this part of the discussion, we believe that topological constraints constitute an *essential* feature to be retained in *minimal* models. These examples illustrate, that the discussion of the origin of the generic structure is far from academic. Instead a quantitative understanding of the interaction free "null model" is essential for attempts to reconstruct or predict the three dimensional structure [112, 113] or the dynamics of entire cell nuclei. Given an initial conformation of chromosomes (*e.g.*, Rabl-like in *Drosophila*), and any other known large-scale geometrical "static" constraint (*e.g.*, shape of the nucleus, anchoring of centromeres [112], ...) and, given a proper mapping of the simulation *vs.* real time, such "null-models" are likely to provide a description of the large-scale structure and dynamics of nuclear compartmentalization.

Second, regarding the *small-scale* structure ($\lesssim 1$ Mbp) structure of chromosomes, we stress once again that our experimentally-motivated working hypothesis that 3D chromatin organization is driven by short-range specific interactions between genomic regions sharing the same chromatin state has mainly been quantitatively investigated in *Drosophila*. As a matter of fact, it is still questionable in higher vertebrates like mammals. On this point, it has been reported recently that about half of the TADs in mammals contain strong loops between oriented CTCF sites usually located at the two boundaries of the domain [43]. While the formation of such loops can still be explained using the same class of models [83, 90, 94], the pivotal observation that looping mainly occurs between convergent CTCF sites [43] is incompatible with short-range interactions [90, 92]. Recently, it was shown that such observations are consistent with an active extrusion mechanism [90, 92]. Protein complexes, putatively cohesins or condensins, bind to chromatin and extrude sequentially large DNA loops before eventually unbinding or stopping at specific loci like CTCF sites having the proper orientation. This model suggests that the local 3D organization is controlled by the presence and orientation of 1D barriers. Polymer models implementing this mechanism have shown that TAD formation and loop interaction at the corners of the domains could be explained by the extrusion process. Moreover, such models can quantitatively predict the perturbed 3D organization after deletion, inversion or duplication of CTCF

sites [39, 90, 114, 115]. They also provide a very elegant mechanism for the formation of mitotic chromosomes and for the separation of sister chromatids, arising from an increase in the number of loop extruders coupled to a decrease in the number of boundary elements [116, 117]. However, loop extrusion cannot account for long-range communications between TADs, for the formation of the A/B compartments or for interactions with the nuclear membranes that are likely to be driven by genomic or chromatin-associated information. Heteropolymer models accounting for both loop extrusion and specific short-range interactions remain to be developed in order to quantitatively describe within the same framework the local and higher-order chromosome organization in mammals.

Interestingly, a still open question is if the spatial organization of chromatin resulting (in part) from the clustering of chromatin states is only a *by-product* of genome activity or is actively participating to the local regulation of the chromatin assembly and more generally to the regulation of the genome function. An attractive hypothesis is that 3D domains (TADs, A/B compartments) would correspond to nano-reactors: a few number of chromatin-associated complexes co-localize in space, increasing their local concentration and thus promoting their biochemical activity on chromatin. Nucleation by a small number of factors coupled to self-assembly or multimerization of biomolecules leads to the formation of interaction domains which further enhance, stabilize and/or perpetuate the active or repressed environment. TADs would correspond to sub-reactors, having a role in either preventing or facilitating the communication between distal regulatory genomic elements at the sub-Mbp scale thus enhancing efficiency of genes co-activation or co-repressions [41, 118]. Domain sizes through the control of global compaction may have co-evolved in order to increase the robustness of these regulatory contacts, for example to motif mutations [41]. The (self-)assembly of TADs into A/B compartments is a softer mode of regulation where spatial confinement increase binding affinities to the regulated sequences. Development of mixed models coupling the heteropolymer description to standard gene or epigenetic regulation dynamics [119–121] would certainly be very helpful in the near future to theorize and quantify such concepts but also to interpret more deeply experimental observations.

Acknowledgements – We thank Peter Meister and Giacomo Cavalli for fruitful discussions. We acknowledge our funding agencies: Agence Nationale de la Recherche (ANR-15-CE12-0006 EpiDevoMath; DJ/CV/RE), Fondation pour la Recherche Médicale (DEI20151234396; DJ/CV), CNRS (DJ/CV), Institut Rhône-Alpin des Systèmes Complexes (DJ), program AGIR of University Grenoble Alpes (DJ), ENS de Lyon (CV/RE), Italian Ministry of Education (PRIN 2010HXAW77; AR), SISSA (AR). Computer simulations discussed in this work were performed at Cineca (Bologna, Italy), CBP and PSMN (ENS-Lyon, France) and P2CHPD (UCB Lyon 1, France) by employing in part the equip@meso facilities of the FLMSN.

[1] C. Mora, D. P. Tittensor, S. Adl, A. G. B. Simpson, and B. Worm, Plos Biol. **9**, 1 (2011).
[2] V. A. Lukhtanov, Comp. Cytogenet. **9**, 683 (2015).
[3] I. J. Leitch, D. E. Soltis, P. S. Soltis, and M. D. Bennett, Ann. Botany **95**, 207 (2005).
[4] M. D. Bennett and I. J. Leitch, in *The Evolution of the Genome*, edited by T. R. Gregory (Elsevier, San Diego, 2005), pp. 89–162.
[5] T. R. Gregory, in *The Evolution of the Genome*, edited by T. R. Gregory (Elsevier, San Diego, 2005), pp. 3–87.
[6] http://www.ensembl.org/index.html.
[7] J. Pellicer, M. F. Fay, and I. J. Leitch, Bot. J. Linn. Soc. **164**, 10 (2010).
[8] B. Alberts et al., *Molecular Biology of the Cell* (Garland Science, New York, 2007), 5^{th} ed.
[9] T. Cremer and C. Cremer, Nature Rev. Genet. **2**, 292 (2001).
[10] Y. Weiland, P. Lemmer, and C. Cremer, Chromosome Res. **19**, 5 (2011).
[11] J. Dekker, M. A. Marti-Renom, and L. A. Mirny, Nat. Rev. Genet. **14**, 390 (2013).
[12] S. Shachar, G. Pegoraro, and T. Misteli, Cold Spring Harb. Symp. Quant. Biol. **80**, 73 (2015).
[13] J. Fraser, I. Williamson, W. A. Bickmore, and J. Dostie, Microbiol. Mol. Biol. Rev. **79**, 347 (2015).
[14] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Oxford University Press, New York, 1986).
[15] A. Y. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules* (AIP Press, New York, 1994).
[16] M. Rubinstein and R. H. Colby, *Polymer Physics* (Oxford University Press, New York, 2003).
[17] K. Kremer and G. S. Grest, J. Chem. Phys. **92**, 5057 (1990).
[18] A. Y. Grosberg, Polym. Sci. Ser. C **54**, 1 (2012).
[19] J. D. Halverson, J. Smrek, K. Kremer, and A. Y. Grosberg, Rep. Prog. Phys. **77**, 022601 (2014).
[20] A. Rosa and C. Zimmer, Int. Rev. Cell Mol. Biol. **307**, 275 (2014).
[21] A. Grosberg, Y. Rabin, S. Havlin, and A. Neer, Europhys. Lett. **23**, 373 (1993).
[22] A. Rosa and R. Everaers, PLoS Comput. Biol. **4**, e1000153 (2008).
[23] T. Vettorel, A. Y. Grosberg, and K. Kremer, Phys. Today **62**, 72 (2009).
[24] J. R. Dixon et al., Nature **485**, 376 (2012).
[25] T. Sexton et al., Cell **148**, 458 (2012).
[26] M. Emanuel, N. H. Radja, A. Henriksson, and H. Schiessel, Phys. Biol. **6**, 025008 (2009).
[27] P. Heun, T. Laroche, K. Shimada, P. Furrer, and S. M. Gasser, Science **293**, 2181 (2001).
[28] E. Lieberman-Aiden et al., Science **326**, 289 (2009).
[29] K. Bystricky, P. Heun, L. Gehlen, J. Langowski, and

S. M. Gasser, Proc. Natl. Acad. Sci. USA **101**, 16495 (2004).
[30] R. K. Sachs, G. van den Engh, B. Trask, H. Yokota, and J. E. Hearst, Proc. Natl. Acad. Sci. USA **92**, 2710 (1995).
[31] J. Mateos-Langerak et al., Proc. Natl. Acad. Sci. USA **106**, 3812 (2009).
[32] A. Rosa and R. Everaers, Phys. Rev. Lett. **112**, 118302 (2014).
[33] T. Cremer and C. Cremer, Eur. J. Histochem. **50**, 161 (2006).
[34] C. Rabl, Morphologisches Jahrbuch **10**, 214 (1885).
[35] Some notable exceptions have been described. For instance, detection of territories in organisms with small genomes like yeast has appeared to be elusive [122]. Yet, some "loose" territoriality at the gene level has been reported [123].
[36] P.-G. De Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, 1979).
[37] C. Münkel et al., J. Mol. Biol. **285**, 1053 (1999).
[38] J. W. K. Ho et al., Nature **512**, 449 (2014).
[39] E. P. Nora et al., Nature **485**, 381 (2012).
[40] J. R. Dixon et al., Nature **485**, 376 (2012).
[41] T. Sexton and G. Cavalli, Cell **160**, 1049 (2015).
[42] J. R. Dixon et al., Nature **518**, 331 (2015).
[43] S. S. P. Rao et al., Cell **159**, 1665 (2014).
[44] M. Imakaev et al., Nat. Methods **9**, 999 (2012).
[45] J. Fraser et al., Mol. Syst. Biol. **11**, 852 (2015).
[46] Y. Zhu et al., Nat. Commun. **7**, 10812 (2016).
[47] K. Maeshima, S. Hihara, and M. Eltsov, Curr. Opin. Cell Biol. **22**, 291 (2010).
[48] O. Kratky and G. Porod, Rec. Trav. Chim. Pays-Bas. **68**, 1106 (1949).
[49] N. B. Becker, A. Rosa, and R. Everaers, Eur. Phys. J. E **32**, 53 (2010).
[50] The nucleosome core particle (*i.e.* the histone octamer plus the wrapped DNA) has a roughly cylindrical shape with a diameter of 10nm and a height of 6nm [124]. With around 50bp per linker, the typical distance between the centers of neighboring core particles is of the order of "10nm + 50bp/(3bp/nm) = 25nm". The contour length density of the 10nm fiber is hence "200bp/(25nm)=8bp/nm". Allowing for variations in the linker length (a variation of "± one base pair" corresponds to a rotation of ±34 degrees around the linker axis due to the twist of the double-helix) [125], for over- and underwrapping of DNA in the core particle as well as for the conformational flexibility of the linkers, subsequent "bond vectors" between core particle centers are to a first approximation uncorrelated. In this case, the Kuhn length of the 10nm-fiber equals $l_{K,10nm} = 25$nm.
[51] A.-M. Florescu, P. Therizols, and A. Rosa, Plos. Comput. Biol. **12**, e1004987 (2016).
[52] S. F. Edwards, Proc. Phys. Soc. **91**, 513 (1967).
[53] S. Prager and H. L. Frisch, J. Chem. Phys. **46**, 1475 (1967).
[54] T. A. Kavassalis and J. Noolandi, Phys. Rev. Lett. **59**, 2674 (1987).
[55] L. J. Fetters, D. J. Lohse, D. Richter, T. A. Witten, and A. Zirkel, Macromolecules **27**, 4639 (1994).
[56] N. Uchida, G. S. Grest, and R. Everaers, J. Chem. Phys. **128**, 044902 (2008).
[57] G. G. Cabal et al., Nature **441**, 770 (2006).
[58] T. C. B. McLeish, Adv. Phys. **51**, 1379 (2002).
[59] P.-G. De Gennes, J. Chem. Phys. **55**, 572 (1971).
[60] J.-L. Sikorav and G. Jannink, Biophys. J. **66**, 827 (1994).
[61] A. Y. Grosberg, S. K. Nechaev, and E. I. Shakhnovich, J. Phys. France **49**, 2095 (1988).
[62] R. D. Schram, G. T. Barkema, and H. Schiessel, J. Chem. Phys. **138**, 224901 (2013).
[63] A. Rosa, N. B. Becker, and R. Everaers, Biophys. J. **98**, 2410 (2010).
[64] T. Vettorel, A. Y. Grosberg, and K. Kremer, Phys. Biol. **6**, 025013 (2009).
[65] S. P. Obukhov, M. Rubinstein, and T. Duke, Phys. Rev. Lett. **73**, 1263 (1994).
[66] A. R. Khokhlov and S. K. Nechaev, Phys. Lett. **112A**, 156 (1985).
[67] M. E. Cates and J. M. Deutsch, J. Phys. (Paris) **47**, 2121 (1986).
[68] J. Klein, Macromolecules **19**, 105 (1986).
[69] M. Rubinstein, Phys. Rev. Lett. **57**, 3023 (1986).
[70] M. G. Brereton and T. A. Vilgis, J. Phys. A: Math. Gen. **28**, 1149 (1995).
[71] M. Müller, J. P. Wittmer, and M. E. Cates, Phys. Rev. E **53**, 5063 (1996).
[72] M. Müller, J. P. Wittmer, and M. E. Cates, Phys. Rev. E **61**, 4078 (2000).
[73] J. Suzuki, A. Takano, T. Deguchi, and Y. Matsushita, J. Chem. Phys. **131**, 144902 (2009).
[74] J. D. Halverson, W. B. Lee, G. S. Grest, A. Y. Grosberg, and K. Kremer, J. Chem. Phys. **134**, 204904 (2011).
[75] A. Y. Grosberg, Soft Matter **10**, 560 (2014).
[76] A. Rosa and R. Everaers, Submitted (2016).
[77] A. Khalil et al., Chromosome Res. **15**, 899 (2007).
[78] H. H. Q. Heng et al., Proc. Natl. Acad. Sci. USA **93**, 2795 (1996).
[79] A. Sanyal, B. R. Lajoie, G. Jain, and J. Dekker, Nature **489**, 109 (2012).
[80] M. Bohn and D. W. Heermann, Plos One **5**, e12218 (2010).
[81] M. Barbieri et al., Proc. Natl. Acad. Sci. USA **109**, 16173 (2012).
[82] H. Jerabek and D. W. Heermann, PLoS One **7**, e37525 (2012).
[83] C. A. Brackley, J. Johnson, S. Kelly, P. R. Cook, and D. Marenduzzo, Nucleic Acids Res. **44**, 3503 (2016).
[84] F. Benedetti, J. Dorier, Y. Burnier, and A. Stasiak, Nucleic Acids Res. **42**, 2848 (2014).
[85] B. Doyle, G. Fudenberg, M. Imakaev, and L. A. Mirny, Plos Comput. Biol. **10**, e1003867 (2014).
[86] N. Ganai, S. Sengupta, and G. I. Menon, Nucleic Acids Res. **42**, 4145 (2014).
[87] D. Jost, P. Carrivain, G. Cavalli, and C. Vaillant, Nucleic Acids Res. **42**, 9553 (2014).
[88] M. Tark-Dame, H. Jerabek, E. M. M. Manders, D. W. Heermann, and R. van Driel, PLoS Comput. Biol. **10**, e1003877 (2014).
[89] L. I. Nazarov, M. V. Tamm, V. A. Avetisov, and S. K. Nechaev, Soft Matter **11**, 1019 (2015).
[90] A. L. Sanborn et al., Proc. Natl. Acad. Sci. USA **112**, E6456 (2015).
[91] S. V. Ulianov et al., Genome Res. **26**, 70 (2016).
[92] G. Fudenberg et al., Cell Rep. **15**, 2038 (2016).
[93] G. Tiana et al., Biophys. J. **110**, 1234 (2016).
[94] A. M. Chiariello, C. Annunziatella, S. Bianco, A. Esposito, and M. Nicodemi, Sci. Rep. **6**, 29775 (2016).

[95] F. S. Bates and G. H. Fredrickson, Annu. Rev. Phys. Chem. **41**, 525 (1990).
[96] D. Canzio et al., Nature **496**, 377 (2013).
[97] K. Isono et al., Dev. Cell **26**, 565 (2013).
[98] K. Hiragami-Hamada et al., Nat. Commun. **7**, 11310 (2016).
[99] C. Allis, T. Jenuwein, and D. Reinberg, *Epigenetics* (Cold Spring Harbor Laboratory Press, 2007).
[100] S. C. Weber, A. J. Spakowitz, and J. A. Theriot, Proc. Natl. Acad. Sci. USA **109**, 7338 (2012).
[101] J. D. Olarte-Plata, N. Haddad, C. Vaillant, and D. Jost, Phys. Biol. **13**, 026001 (2016).
[102] T. Ramalho, M. Selig, U. Gerland, and T. A. Ensslin, Phys. Rev. E **87**, 022719 (2013).
[103] T. Cheutin and G. Cavalli, Plos Genet. **8**, e1002465 (2012).
[104] N. Naumova et al., Science **342**, 948 (2013).
[105] A. N. Boettiger et al., Nature **529**, 418 (2016).
[106] S. Sofueva et al., Embo J. **32**, 3119 (2013).
[107] T. Pederson, Mol. Biol. Cell **11**, 799 (2000).
[108] P. R. Cook, Science **284**, 1790 (1999).
[109] P. R. Cook, J. Mol. Biol. **395**, 1 (2010).
[110] F. J. Nedelec, T. Surrey, A. C. Maggs, and S. Leibler, Nature **389**, 305 (1997).
[111] P. W. K. Rothemund, Nature **440**, 297 (2006).
[112] H. Wong et al., Curr. Biol. **22**, 1881 (2012).
[113] D. Baù et al., Nat. Struct. Mol. Biol. **18**, 107 (2011).
[114] D. G. Lupiáñez et al., Cell **161**, 1012 (2015).
[115] Y. Guo et al., Cell **162**, 900 (2015).
[116] A. Goloborodko, M. V. Imakaev, J. F. Marko, and L. Mirny, Elife **5** (2016).
[117] A. Goloborodko, J. F. Marko, and L. A. Mirny, Biophys J. **110**, 2162 (2016).
[118] B. Tolhuis et al., Plos Genet. **7**, e1001343 (2011).
[119] D. J. Wilkinson, Nat. Rev. Genet. **10**, 122 (2009).
[120] I. B. Dodd, M. A. Micheelsen, K. Sneppen, and G. Thon, Cell **129**, 813 (2007).
[121] D. Jost, Phys. Rev. E **89**, 010701 (2014).
[122] J. E. Haber and W. Y. Leung, Proc. Natl. Acad. Sci. USA **93**, 13949 (1996).
[123] A. B. Berger et al., Nat. Methods **5**, 1031 (2008).
[124] K. Luger and J. C. Hansen, Curr. Opin. Struct. Biol. **15**, 188 (2005).
[125] J. Yao, P. T. Lowary, and J. Widom, Proc. Natl. Acad. Sci. USA **87**, 7603 (1990).